Am. J. Hum. Genet. 53:1-5, 1993

1992 William Allan Award Address

Alec J. Jeffreys

Department of Genetics, University of Leicester, Leicester

I am deeply honored by the Allan Award, for myself and for all my colleagues who have been involved in the often labyrinthine pathway of research that led initially from a freezer at the British Antarctic Survey, Cambridge, and culminated eventually in DNA fingerprinting. One of the great pleasures of science for me is the unexpected and serendipitous twists of fate that can shift the emphasis of research programs and, indeed, entire careers. Such was the story of DNA fingerprinting; I think that I would have been as astonished at the suggestion, 15 years ago, that I would one day be involved in using DNA in criminal investigations, as I am today at receiving the Allan Award.

My story starts in 1975 when, as a young and naive postdoc, I joined Dick Flavell in Amsterdam on a joint project with Charles Weissmann to isolate the rabbit β-globin gene by a combination of biochemical purification and cloning. During this project, we managed to develop Southern blot hybridization to the point where we could detect gene fragments not only in our partially purified fractions but also, and astonishingly at the time, in total genomic DNA—our first glimpse of a mammalian single-copy gene. Restriction analysis soon showed that the gene was a discrete entity amenable to physical mapping and was distinctly odd, with a gap in the middle—one of the first examples of an intron (Jeffreys and Flavell 1977).

In the summer of 1977, I moved to the Department of Genetics at Leicester as a junior lecturer and was faced with the daunting prospect of establishing an independent research program with just myself and a part-time technician who had barely heard of DNA. This was the time for hard decisions. Should I press on with the intron work, despite the entry of many large groups into the field? Should I move in another, rather obvious, direction, using globin gene probing to hunt for molecular defects in the hemoglobinopathies? Instead, I decided on a major shift in direction, towards tackling problems of DNA variation and the evolution of gene families, an almost virgin field. As a graduate student at the Genetics Laboratory, Oxford, I had been raised on a staple diet of biochemical and serological polymorphisms, and I appreciated the limitations of these systems, particularly their restricted number and the extremely heterogeneous mode of detection. I reasoned that, if we could construct restriction maps around genes, then, with luck, we should be able to detect DNA polymorphisms that affected restriction sites, in other words, RFLPs. By late 1977 we had our first variant, a rare δ-globin RFLP found in my technician. To verify its heritability, we tested her parents—and so, I suspect, were the first subconscious seeds of paternity analysis sown in my mind.

This early survey of the incidence of RFLPs in human DNA (Jeffreys 1979), together with the work of Kan and Dozy (1978), showed that RFLPs were likely to be reasonably common in human DNA and could provide a unified approach to developing unlimited numbers of human genetic markers and thereby solve one of the major hurdles in the construction of human linkage maps. Shortly thereafter, this idea was greatly elaborated by Botstein, White, Skolnick, and Davis (Botstein et al. 1980), who defined the strategy and feasibility of global linkage mapping using RFLPs. This strategy has come to recent spectacular fruition through a major international collaboration (NIH/CEPH Collaborative Mapping Group 1992).

By the early 1980s, it was clear that RFLPs were not only rather difficult to find but also cumbersome and expensive to type, as well as being distinctly uninformative (almost all were diallelic systems). Would it be possible to find much more variable and informative loci? To my delight, Arlene Wyman and Ray White (Wyman and White 1980) reported the fortuitous isolation of the first hypervariable multiallelic locus, though, at the time, the physical basis of variability was unclear. It was

Received December 11, 1992. This address was delivered November 13, 1992, in San Francisco.

Address for correspondence and reprints: Alec J. Jeffreys, Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom.

0002-9297/93/5301-0001$02.00

with great interest, therefore, that I read the paper of Bell et al. (1982), who showed that another hypervariable locus found near the human insulin gene consisted of tandem repeats, with multiallelic variation arising through variation in repeat copy number. These multiallelic loci seemed to be the key which would unlock the door to efficient linkage analysis in man. In 1982, we therefore turned our thoughts to developing general methods for detecting and isolating such loci. Various harebrained schemes, such as cloning regions of DNA devoid of restriction sites, were tried and rejected. As so often happens in science, the answer came from a wholly different and unexpected direction.

Since 1978, we had been pursuing a research program studying the molecular evolution of gene families, particularly the globin genes. Paul Simon, a perceptive young undergraduate student, pointed out that our globin-gene family studies were incomplete, since no one had yet isolated the myoglobin gene, a distant relative of this family. We therefore cloned the cDNA and gene via seal skeletal muscle, a rich source of myoglobin mRNA. (The seal muscle was kindly provided by the British Antarctic Survey, who keep a collection of tissue samples of marine mammals—hence our story starting in a British Antarctic Survey freezer). Routine sequence analysis of the human myoglobin gene revealed a short tandem repeat region in one of the introns, a region we were later to dub a "minisatellite" by virtue of its structural similarity, on a small scale, to classical satellite DNA. Curiously, there were sequence similarities between this minisatellite and a few other known minisatellites (Weller et al. 1984); others too had noted that minisatellite sequences did not appear to be random (Proudfoot et al. 1982; Goodbourn et al. 1983). Was there a sequence motif shared by minisatellites, and, if so, could this provide the basis of a general hybridization probe for detecting hypervariable loci?

In 1983, Vicky Wilson and I started the quest by searching for genomic clones which could (just about) cross-hybridize to the myoglobin minisatellite. To our delight, all clones isolated contained a minisatellite, some of which were polymorphic in human DNA. More intriguingly, sequence analysis revealed a clear 10–15-bp common "core" sequence embedded within the repeat units of each locus (Jeffreys et al. 1985*a*). Two ideas immediately sprang to mind: first, that this core sequence might in some way be promoting the generation of hypervariable loci, and, second, that probes containing tandem repeats of the core might be particularly effective at detecting many minisatellites simultaneously. Fortunately, two of our clones consisted of tandem repeats of just the core sequence, and so we tried the obvious experiment—namely, Southern blot hybridization of genomic DNA with these probes. The results, developed early one Monday morning in September 1984, instantly and literally changed my life. The probes had managed to detect an astonishingly rich, complex, and variable array of DNA fragments, not just from man but also from several other animal species that happened to be on the Southern blot. Unknowingly, we had stumbled upon DNA fingerprinting (Jeffreys et al. 1985*a*, 1985*b*). The implications for individual identification and kinship analysis were obvious. We knew that conventional serological and biochemical polymorphisms had a distinguished history of forensic service dating back to the first suggestion by Landsteiner and Richter (1903) that ABO typing could be used for bloodstain identification. However, we also knew that these markers were of limited utility and could normally only be used in an exclusionary context. It was clear that these hypervariable DNA patterns offered the promise of a truly individual-specific identification system. We therefore coined the term "DNA fingerprinting" as a deliberate move to emphasize the new forensic paradigm that we could foresee if these probes could be used in criminal and civil investigations.

Our next job was to explore the genetic properties of these DNA fingerprints, including individual variability, the nature and genomic distribution of the hypervariable loci contributing to a DNA fingerprint, and the mutational properties of these unstable loci. All of these studies confirmed our view of the extraordinary levels of individual specificity (except for monozygous twins) afforded by DNA fingerprinting, and its potential utility in definitively assigning or excluding parentage in paternity disputes. Peter Gill at the Home Office Forensic Science Service and I also collaborated to show that DNA fingerprinting could be applied to at least some forensic samples, such as blood and semen stains (Gill et al. 1985). At the time, however, I felt that a radical new technology like this would take years of validation before ever becoming acceptable in court. I had not, however, taken into the account the enthusiasm of English lawyers!

In April 1985, I was contacted by Sheona York of the Hammersmith and Fulham Community Law Centre. She had heard of our work and wondered whether DNA fingerprinting could be used to help an immigrant family whose youngest son was facing deportation on the grounds of strong circumstantial evidence suggesting that the boy was bogus and not a true member of

the family. Blood-group evidence supported the boy's claim, but not convincingly. With some trepidation, we tackled the case and showed that, beyond any conceivable doubt, the boy was a full authentic member of the family (Jeffreys et al. 1985*c*). We submitted the evidence to the Home Office, who obviously had never heard of DNA, yet alone DNA fingerprinting. The authorities nevertheless promptly dropped the case against the boy. The ensuing publicity opened a floodgate of enquiries from immigrant communities in which thousands of families, particularly from the Indian subcontinent, were separated by immigration disputes over claimed family relationships. By early 1986, we had also had our first paternity case accepted in a magistrates court, and we soon found ourselves in a state of siege. To solve the demand, Cellmark Diagnostics was created to provide the first commercial DNA fingerprinting service; to date, they have processed over 90,000 samples for paternity and immigration analysis and have made a massive impact on civil law and immigration policy in the United Kingdom.

Multilocus DNA fingerprinting, using either the initial minisatellite probes or a host of alternative probes based on synthetic or natural tandem repeats, has now become standard practice in resolving kinship disputes, at least in the United Kingdom. This technology has also found major applications in animal and plant analysis, particularly for investigating population structure and reproductive strategies and for conservation biology. The complexity of DNA fingerprints, their limited sensitivity, and their vulnerability to DNA degradation have, however, severely limited their usefulness in forensic casework.

By mid 1985, we and others set about using these multilocus probes to isolate individual minisatellites, both to study the biology of these loci and to develop highly informative DNA markers for linkage mapping and DNA typing. Yusuke Nakamura and Ray White developed a cosmid screening strategy which yielded many minisatellites, now also known as VNTRs (Nakamura et al. 1987). We followed a different strategy, selecting the longest and most variable alleles for cloning to yield some of the most variable loci ever discovered in human DNA (Wong et al. 1987). The highly variable two-allele hybridization patterns obtained were soon shown to have potential forensic applications, particularly in view of probe sensitivity, profile simplicity, and the ability to database patterns as estimated DNA fragment lengths. These patterns, *per probe,* are not individual specific, and we term them "DNA profiles"; it is with some frustration that I note that the term "DNA fingerprinting" has become corrupted, particularly in the United States, to refer to almost any DNA typing system—I prefer to restrict the use of "DNA fingerprinting" specifically to the individual-specific multilocus patterns generated by minisatellite core probes, as originally defined.

DNA profiling saw its forensic debut in late 1986, with the Enderby murder case local to Leicester, in which two schoolgirls had been raped and murdered. A youth had been arrested and had confessed to one of the murders. The local police asked if I would analyze semen samples from both victims to verify his guilt—if possible, for both murders. The DNA profile evidence obtained was astonishing: the semen from both victims was indeed almost certainly from the same man, but totally mismatched the young man. Following independent testing by Home Office forensic scientists, the police accepted that the confession was false and released the young man, the first person proved innocent by DNA typing. There followed the first-ever DNA-based manhunt, which involved mass voluntary screening of the local population and led to the eventual entrapment of the double murderer, who is now serving two life sentences for his crimes (Wong et al. 1987; Wambaugh 1989).

In the following years, DNA profiling has spread to hundreds of state and commercial forensic laboratories worldwide and is now a standard tool in the forensic scientists' armamentarium. It is unlikely to have escaped the attention of the reader that the forensic implementation of DNA profiling has not occurred without controversy. Following the debacle over the 1989 Castro case in New York, there has emerged, particularly in the United States, an increasingly polarized and rancorous debate, which, if nothing else, has served to highlight the lack of congruence between scientific and legal logic. Every aspect of DNA profiling has come under the microscope, including the technical competence of testing laboratories, the criteria used for declaring profile matches in the face of experimental errors in DNA fragment length measurement, and the potential influences of population structuring and inbreeding on the statistical estimation of match frequencies. The latter area is now the focus of most controversy. While I doubt whether anything I write here would influence the debate one jot, I would like to point out that extensive empirical studies so far have repeatedly failed to reveal population effects of a magnitude sufficient to lead to any likelihood of a miscarriage of justice. I would also note that the novel population-genetic behavior of minisatellites, with their high

mutation rate and demonstrable propensity towards convergent allele length evolution, sets them aside from conventional loci and can make them especially refractory to genetic drift processes.

The next revolution in forensic DNA analysis came in 1988, with Cetus' introduction of user-friendly PCR driven by *Taq* polymerase (Saiki et al. 1988) and the promise of supersensitive typing applicable to the 40% or so of crime-scene samples with too little or too degraded DNA for DNA profiling. We soon showed that minisatellites could be amplified, both singly and in multilocus format, by multiplex PCR, and typed even at the single-molecule or single-cell level (Jeffreys et al. 1988*b*). Additional PCR-based typing methods also appeared, including HLA-DQα typing by reverse dot blot, mtDNA control region typing, and the development of short variable tandem-repetitive DNA markers (CA repeat loci, microsatellites, and simple tandem-repeat loci). The latter in particular offer great forensic promise (Edwards et al. 1991) and have already been applied to difficult casework, such as bone DNA analysis (Hagelberg et al. 1991), and most recently to the successful identification of the skeletal remains of Josef Mengele, the notorious Auschwitz camp doctor (Jeffreys et al. 1992).

Microsatellites, while effective, do have some limitations, including restricted variability (typically 12 or so alleles per locus) and the need to define alleles by length measurement. In 1990, we therefore sought to develop an alternative DNA typing system which would combine the variability of minisatellites with the speed and sensitivity of PCR and which would further avoid the problems inherent in DNA fragment length measurement. To this end, we have developed a technique called "MVR-PCR" (minisatellite variant repeat mapping by PCR; Jeffreys et al. 1991), which taps into a second level of minisatellite variability—namely, variation in the interspersion pattern of subtly different types of repeat unit along the tandem array. For example, if a locus contains two types of repeat unit, A and B (which might differ by just a single base substitution), then the structure of a given allele could be represented by AABABBABABAAA . . .—in other words, a binary code. For total genomic DNA, both alleles are superimposed to give a ternary code 12331312133322 . . . , where 1 = both repeats A type at a given position, 2 = both B type, and 3 = heterozygous A/B. We have shown that MVR-PCR can generate extraordinarily informative unambiguous digital codes from human DNA (500 unrelated people so far typed all have different codes at just a *single* locus) and circumvents many of the limitations of conventional DNA typing. We are currently evaluating its applicability to forensic casework, with encouraging results so far. MVR-PCR has also revealed, for the first time, the true level of allelic variability at a minisatellite locus, with at least 10^8 different alleles at the MS32 locus in the world today, compared with a mere 100 or so alleles distinguishable by allele length.

In parallel with the forensic applications of minisatellites, we have continued exploring their biological properties. In 1985, we first suggested that the minisatellite core sequence may in some way promote instability, perhaps by serving as a recombination signal to drive unequal exchange between minisatellite alleles (Jeffreys et al. 1985*a*). We were particularly intrigued when Nicola Royle in my laboratory first showed that minisatellites are not dispersed at random in the human genome but cluster near telomeres (Royle et al. 1988). While limiting their usefulness in linkage analysis, this clustering in or near genomic regions known from cytological studies to be involved in homologue recognition at meiosis, synapsis, and recombination suggested some possible involvement of minisatellites in these chromosomal processes. Direct detection of new mutant alleles at these unstable loci, both by pedigree analysis and by single-molecule PCR analysis of germ-line and somatic DNA (Jeffreys et al. 1988*a*, 1990, 1991), has allowed us to analyze the internal structure of new mutant alleles directly. It is now clear that minisatellites (at least those we have analyzed) do not mutate solely by replication slippage, the mechanism believed to operate at microsatellite loci. Instead, clear evidence is emerging for recombinational interplay between sister chromatids and homologous chromosomes in the mutation process, consistent with a role for minisatellites in recombination. Curiously, many events are directed towards the extreme beginning of the tandem-repeat array creating a localized variability and mutation hot spot. Some mutation events have the characteristics of localized gene conversions at this hot spot, consistent, for example, with gap repair of a double-stranded break introduced into the beginning of the tandem array of the recipient allele. Localized mutation thus suggests a recombination/conversion hot spot present at the beginning of the array; further definition of the role of this hot spot in minisatellite instability is being pursued by a search for *trans*-acting minisatellite DNA-binding proteins and by studying human minisatellites in transgenic mice to define *cis*-acting flanking sequences which serve as recombinators. It remains to be seen whether minisatellite are just another category of selfish, though very

useful, DNA or instead exert some profound influence on fundamental chromosomal processes.

Finally, I should like to express my gratitude to all my colleagues, both within and without science, for their unstinting support over an exciting—and frequently frantic—decade of research. It has been a great privilege to be involved in developing a field with such obvious social, legal, and political impact. One of the highest compliments can be found in the 8th edition of the *Concise Oxford Dictionary*, which includes the entry "genetic fingerprinting (or profiling), the analysis of characteristic patterns in DNA as a means for identifying individuals"; this establishes, if nothing else, that DNA typing is now part of the English language!

References

Bell GI, Selby MJ, Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31–35

Botstein D, White RL, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Gen 32:314–331

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic applications of DNA 'fingerprints'. Nature 318:577–579

Goodbourn SEY, Higgs DR, Clegg JB, Weatherall DJ (1983) Molecular basis of length polymorphism in the human ζ-globin gene complex. Proc Natl Acad Sci USA 80:5022–5026

Hagelberg E, Gray IC, Jeffreys AJ (1991) Identification of the skeletal remains of a murder victim by DNA analysis. Nature 352:427–429

Jeffreys AJ (1979) DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes in man. Cell 18:1–10

Jeffreys AJ, Allen MJ, Hagelberg E, Sonnberg A (1992) Identification of the skeletal remains of Josef Mengele by DNA analysis. Forensic Sci Int 56:65–76

Jeffreys AJ, Brookfield JFY, Semeonoff R (1985*c*) Positive identification of an immigration test-case using human DNA fingerprints. Nature 317:818–819

Jeffreys AJ, Flavell RA (1977) The rabbit β-globin gene contains a large insert in the coding sequence. Cell 12:1097–1108

Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204–209

Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying allelic variation and mutation by single molecule analysis. Cell 60:473–485

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988*a*) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332:278–281

Jeffreys AJ, Wilson V, Neumann R, Keyte J (1988*b*) Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Res 16:10953–10971

Jeffreys AJ, Wilson V, Thein SL (1985*a*) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–74

———(1985*b*) Individual-specific 'fingerprints' of human DNA. Nature 316:76–79

Kan YW, Dozy AM (1978) Polymorphism of DNA sequence adjacent to human β-globin structure gene: relationship to sickle mutation. Proc Natl Acad Sci USA 75:5631–5635

Landsteiner K, Richter M (1903) Über die Verwertbarkeit individueller Blutdifferenzen für die forensische Praxis. Med Beamte 16:85–89

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, et al (1987) Variable numbers of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–1622

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:67–86

Proudfoot NJ, Gil A, Maniatis T (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical, pseudogene. Cell 31:553–563

Royle NJ, Clarkson RE, Wong Z, Jeffreys AJ (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics 3:352–360

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Wambaugh J (1989) The blooding. William Morrow, New York

Weller P, Jeffreys AJ, Wilson V, Blanchetot A (1984) Organization of the human myoglobin gene. EMBO J 3:439–446

Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann Hum Genet 51:269–288

Wyman A, White R (1980) A highly polymorphic locus in human DNA. Proc Natl Acad Sci USA 77:6754–6758